AI-based prediction of crop yield using environmental conditions and data sets

A Proposal submitted to

SWAMI RAMANAND TEERTH MARATHWADA UNIVERSITY,

Nanded

In

Computer Sciences

Under the School of Computational Sciences

By

Mr. Abdulhefedh Mohammed Ali Alsarfi

UNDER THE GUIDANCE OF

Dr. Archana H. Sable

2025-2026

Abstract

This proposal will examine the use of deep learning technologies to predict agricultural diseases and pests based on environmental data such as soil properties, weather, crop production, as well as soil moisture data from SMAP satellites. Previous studies show that algorithms such as Random Forest and sgdclassifier achieve high accuracy in crop recommendation, while CNN technologies help detect plant diseases via images. However, limited data, poor generalization between regions, and limited reliance on remote sensing data remain major challenges. The proposal suggests an integrated system that combines multiple data sources with deep learning and machine learning models, as well as an Android application that provides farmers with instant recommendations on: the most suitable crop, the best time for planting and harvesting, and alerting of the presence of diseases or pests. The results confirm that integrating AI with environmental data can improve crop productivity and contribute to smart agriculture.

1. Introduction

Farming is the primary and most important occupation in India. Almost 50% of India's manpower has their occupation directly or indirectly linked to agriculture. India is hence rightfully called as an agricultural country. Agriculture provides almost 20% contributions to the Gross Domestic Product. India has exported agricultural goods worth $25 billion until November 2017.[1] Productivity needs to be increased so that farmers can get more pay from the same piece of land with less labour. Precision agriculture provides a way to do it. Precision farming, as the name implies, refers to the applying of precise and proper total of comment-like pee, fertilizers, soil etc. at the proper time to the craw for increasing its productivity and increasing its yields. Not all precision agriculture systems offer best results. But in agriculture it is important that the recommendations made are accurate and precise because in case of errors it may lead to heavy material and capital loss.[2] Sugarcane is one of the multiproduct cash crop and grown in both tropical and subtropical regions in India. In sub-tropical region, Uttar Pradesh, Uttarakhand, Haryana, Punjab, Bihar and in tropical region, Maharashtra, Karnataka, Tamil Nadu, Andhra Pradesh, Gujarat, Madhya Pradesh is the main sugarcane growing states. Sugarcane is pro vides assured returns to the farmers and socio financial improvement of the farming community.[3]

Assessing the soil quality, reviewing agricultural crop policies and single crop recommendations, and understanding the impact of climate anomalies on crops. However, the scope of the existing works reported in the literature is often narrowly defined, either addressing a single aspect of farming or not integrating crucial components of fertilizer suggestion and soil conditions. This isolation often results in a lack of comprehensive insights, limiting the existing approaches' real-world applicability and accuracy.[4]

2. Background

Agriculture plays a critical role in the economy of many countries, including India, where nearly 50% of the workforce is engaged directly or indirectly in farming. Crop productivity is closely tied to soil health, climate conditions, and timely agricultural practices. Traditional farming methods, which rely heavily on manual monitoring and local reporting, often lead to delayed interventions, lower yields, and economic losses. Precision agriculture has emerged as a promising approach to optimize resource utilization such as fertilizers, water, and laborby applying precise recommendations at the right time. However, the effectiveness of precision agriculture depends on the accuracy of the data and the models used to make decisions, and errors in recommendations can result in significant material and financial losses. Recent advancements in artificial intelligence, particularly machine learning (ML) and deep learning (DL), have enabled the development of intelligent systems that can analyze complex environmental data to provide actionable insights for farmers. ML algorithms such as Random Forest, Decision Trees, XGBoost, and SGDC have demonstrated high accuracy in predicting suitable crops based on soil properties, climatic factors, and historical crop data. Similarly, Convolutional Neural Networks (CNNs) have shown promise in detecting plant diseases through image analysis. These technologies offer the potential to improve decision-making in agriculture, increase productivity, and promote sustainable farming practices. To address these challenges, integrated AI-based systems are needed that combine multiple data sources including soil properties, weather conditions, crop production records, and satellite-derived soil moisture measurements with advanced ML and DL models. Such systems can provide farmers with accurate and timely recommendations regarding the most suitable crops, optimal planting and harvesting periods, and early detection of diseases and pests. By leveraging multisource environmental data and state-of-the-art AI techniques, these systems aim to enhance crop productivity, reduce economic and environmental risks, and support the broader goals of smart and sustainable agriculture.

3. Related Work

M. K. Senapaty et al [5], proposed a model for developing a crop recommendation system, primarily leveraging machine learning classification algorithms using the Historical Crop Dataset and Local Agricultural Data. Thirteen different ML classification algorithms were applied; SGDC showed the best performance with an accuracy of 1.00, precision of 1.00, and specificity of 0.54 after SMOTE. The paper explicitly states that a limitation of this research is the non-utilization of sensors and drone technology.

U. R. N. Dagada Madhu [6] This study uses data collected from an official government of India website for the Kurnool region of Andhra Pradesh, and includes 67,788 records processed into 5,649 rows containing 12 soil properties including nitrogen, phosphorus, potassium, PH, electrical conductivity, organic carbon and microelements such as iron, zinc, manganese, sulfur, etc., with the aim of recommending only three crops-rice, cotton and neighbor. Six

machine learning models were applied: Naïve Bayes, logistic regression, SVM, decision tree, random forest, and xgboost, and the results showed the superiority of tree models over non-tree models, where the decision tree achieved an accuracy of 72% and the random forest 83%, while XGBoost clearly outperformed, achieving the highest accuracy of 93.2% and the highest F1 value of 91.6% as a result of its reliance on gradually reducing errors through successive trees and using the first and second derivatives to effectively reduce loss. However, the study suffers from several limitations, the most important of which are its reliance on data from only one geographical area, limiting recommendations to three crops, its full dependence on soil characteristics without introducing climatic factors such as rainfall, humidity and temperature, in addition to not documenting any accurate optimization of coefficients (Hyperparameter Tuning) and not testing the ability of models to generalize to new regions or data, which makes its results strong within its scope but limited in its applicability on a larger scale.

S. K. Upadhyay et al [7] This study was based on data comprising 67,788 records of soil characteristics in the Kurnool region with three crops: rice, cotton and jowar. The data included 12 nutrients such as N, p, k, ph, PEC and others. Six machine learning models have been applied: Naïve bays, Logistic Regression, SVM, Decision Tree, Random Forest, and XGBoost. The results showed the superiority of tree models, with Decision Tree achieving about 72% and random Forest about 83%, while XGBoost provided the highest accuracy of 93.2% and the best F1 at 91.6%. Non-tree models performed less well due to their reliance on assumptions that did not fit the nature of the data. He highlighted the limitations that the data is from only one region, the lack of inclusion of climatic factors, in addition to the sensitivity of some models to the balance in the data. However, the study demonstrated the effectiveness of machine learning in supporting crop selection decisions.

S. K. Upadhyay et al [8] ,Intelligent crop recommendation systems using machine learning rely on various models such as neural networks, decision trees, random forests, conveyor support machines, and complex learning techniques, where data related to soil characteristics, weather, and humidity are used to determine the best crops grown in ambient conditions. Higher accuracy in forecasting suitable crops leads to better decisions of farmers and increased economic productivity. These models rely heavily on the availability of high-quality data that may not always be available, especially in rural or developing areas.

H. Mavi et al [9], This study relies on agricultural data that includes soil elements such as nitrogen, phosphorus and potassium, in addition to temperature, humidity, pH and rainfall to build a crop recommendation system. The study has used a number of machine learning models such as decision tree, knife Bayes, logistic regression, SVM, random forests and xgboost to determine the most suitable among them to predict the appropriate crops. The results showed that both Random Forest and xgboost achieved the highest accuracy of 99.31%, which makes the system an effective tool to support farmers and promote precision farming. However, the study faces some limitations, including the dependence of the model on the quality of available

data, its accuracy is affected by different agricultural environments, in addition to the need for a technical structure that may not be available in some agricultural areas.

S. Bhatnagar et al [10], The study focused on developing a crop recommendation system using machine learning algorithms, relying on comprehensive data including soil, climate, past crop productivity and cultivation methods. She used models such as neural networks, decision trees, random forests and support vector machines (SVMs), and trained on historical data to understand the relationship between the environment and crop performance. The models were evaluated using precision, Loopback and F1-score scales, and the results showed their ability to provide accurate recommendations to farmers. The analysis of the interpretability of models helped to clarify the logic of decision-making. The most prominent disadvantages and limitations are the dependence of performance on data quality, the difficulty of interpreting some models, in addition to the limited effectiveness of the system in new environments. The system was practically implemented via a web application that makes it easy for farmers to enter their data and get personalized recommendations. This system contributes to improving crop selection, increasing productivity, and enhancing the sustainability of agricultural resources.

M. Pan et al [11], The study evaluated soil moisture data from the SMAP mission using a VIC model and field data within the United States, with passive (SMAP_P), compound (SMAP_AP), and active (SMAP_A) product analysis. The results showed that the passive product is very accurate and sometimes outperforms the model, while the active product is weak due to radar failures. The accuracy of the products decreases in humid areas and northern forests, and their impact on vegetation is gradual. Integrating SMAP data with the model improves predictions. Disadvantages include radar failures, differing scales, and a lack of field stations in some areas.

G. Singh et al [12], Heat radiation (brightness temperature, TB) from SMAP L-Band radiometer (coarse resolution) + C-Band radar data from RISAT-1 (SAR), for use in “disaggregation” i.e. dismantling / lowering the TB scale to a higher resolution (1 km / 3 km) with the surface-water mask hydromask is based on RISAT-1 data. At high Beats 1 km and 3 km Ubrmse soil moisture recoveries greater than 0.06 M3/m3 occur in “very wet” conditions, such as when there is a flood or waterlogging. Sensitivity to surface water / pinged water: if the presence of dynamic surface water is ignored, the dismantling of TB is incorrect at high accuracy, which leads to significant errors in soil moisture.

M. A. and T. D. Woldekidan Gudelo,[13]The study deals with the development of an improved predictive model for crops in the Gamo region of Ethiopia, using educational and climatic data from the Agricultural Research Center in Arba Munch and the agricultural guidance office in Gamo, which included 6,560 records and 11 characteristics such as soil type, nitrogen, phosphorus and potassium content, soil pH, altitude, temperature, precipitation, and humidity, in addition to the target crop classification (15 types of crops) . The results outperform the

Extraatrees classifier, achieving a test accuracy of 98.25% and an average weight of accuracy and retrieval measures and F1 exceeding 98%. the SHAP analysis also showed that humidity, temperature, soil type, soil pH, rainfall, and altitude were the most influential on the forecasts. One of the limitations of the study is that it was limited to a specific geographical area, and did not include real-time data collection via IoT yet, which limits the immediate national expansion.

K. K. C. Ryoko Araki et al , [14], the researchers used surface soil moisture data from the SMAP satellite for the period 2015-2022 at a scale of 36 km to analyze the decrease in humidity after rains, excluding data affected by rain, snow, and urban areas. They developed a nonlinear soil loss model by adding the nonlinear coefficient θ to compare plant responses to water with conventional linear models, focusing on the water-Limited et phase that dominates in satellite data. The results showed the superiority of the nonlinear model in explaining most of the low humidity events, revealing that global territories often with a convex loss of water ($\theta>1$) reflect a conservative vegetative use, while grassy and non-woody areas show a concave loss reflecting more aggressive strategies. The disadvantages are that the accuracy of the model is affected by the heterogeneity of the landscape, the retrieval interval, and the limitations of satellite data in deserts and agricultural areas, which suggests using local high-frequency data to improve the assessment, so that the model forms a dynamic framework for understanding the responses of plants to changing water availability.

Table 1: Summary of the related studies.

Reference & Year	**Data Used**	**Method Used**	**Results**	**Strengths**	**Weaknesses**
M. K. Senapaty et al, 2024,[5]	Historical Crop Dataset, Local Agricultural Data	13 ML classification algorithms (SGDC performed best)	SGDC: Accuracy 1.00, Precision 1.00, Specificity 0.54 (after SMOTE)	High predictive accuracy; multiple ML algorithms tested	Did not use sensors or drone technology
U. R. N. Dagada Madhu , 2023, [6]	67,788 records from Govt. of India, 12 soil properties	Naïve Bayes, Logistic Regression, SVM, Decision Tree, Random	Decision Tree: 72% accuracy, Random Forest: 83%, XGBoost: 93.2% accuracy, F1: 91.6%	Strong tree-based model performanc e; detailed soil analysis	Limited geographic area; only 3 crops; no climatic factors; no hyperparamet er tuning; limited

		Forest, XGBoost			generalizability
A. Srinidhi et at,2024,[7]	67,788 soil records, 12 nutrients, 3 crops	Naïve Bayes, Logistic Regression, SVM, Decision Tree, Random Forest, XGBoost	Decision Tree: 72%, Random Forest: 83%, XGBoost: 93.2% accuracy, F1: 91.6%	Confirms effectiveness of ML for crop selection	Single-region data; no climatic factors; model sensitivity to data imbalance
S. K. Upadhyay et at ,2024 ,[8]	Soil characteristics, weather, humidity	Neural Networks, Decision Trees, Random Forests, SVM	High forecasting accuracy improves crop decisions	Highlights importance of multiple ML models	Dependent on high-quality data; limited availability in rural areas
H. Mavi et at .2024 , [9]	Soil elements (N, P, K), temperature, humidity, pH, rainfall	Decision Tree, Naïve Bayes, Logistic Regression, SVM, Random Forest, XGBoost	Random Forest & XGBoost: 99.31% accuracy	Effective precision farming tool; robust ML comparison	Accuracy depends on data quality; affected by agricultural environment; requires technical infrastructure
S. Bhatnagar et at,2024,[10]	Soil, climate, past crop productivity, cultivation methods	Neural Networks, Decision Trees, Random Forests, SVM	Accurate recommendations; interpretable model logic	Web application for practical use; supports sustainable agriculture	Dependent on data quality; some models hard to interpret; limited effectiveness in new environments
M. Pan et at ,2016,[11]	SMAP soil moisture data, VIC	Passive, Compound, and Active	Passive product accurate; active product weak;	Combines satellite and model data;	Radar failures; scale differences;

	model, field data (USA)	SMAP product analysis	integration improves prediction	improves soil moisture prediction	limited field stations
G. Singh et at ,2021,[12]	SMAP L-Band radiometer, RISAT-1 C-Band SAR	Disaggregation of brightness temperature to higher resolution	Errors in very wet conditions (>0.06 m^3/m^3)	Uses multi-sensor data; higher resolution predictions	Sensitive to surface water; misestimation in flooded areas
M. A. and T. D. Woldekidan Gudelo,2025,[13]	6,560 records, 11 features (soil type, nutrients, climate), 15 crops	Extra Trees classifier, SHAP analysis	Test accuracy: 98.25%; F1 >98%	Highlights feature importance; high accuracy	Limited to one region; no real-time IoT data; limited national scalability
K. K. C. Ryoko Araki, Bryn E. Morgan, Hilary K. McMillan,2025,[14]	SMAP surface soil moisture 2015–2022, satellite data	Nonlinear soil loss model (θ coefficient)	Nonlinear model explains low humidity events; global vegetative patterns identified	Novel nonlinear approach; insight into plant-water response	Accuracy affected by landscape heterogeneity, satellite limitations; local high-frequency data needed

4. Problem statement

Farmers face significant challenges in the early detection of agricultural diseases and pests due to limited environmental data and the inefficiency of traditional systems in predicting agricultural risks. Traditional methods are often based on manual monitoring and local reporting, which leads to significant delays in making important agricultural decisions, and, consequently, lower productivity and increased losses. In addition, the available data is often insufficient, not uniform between different regions, which reduces the ability of traditional models to accurately generalize to new regions.

The problem becomes more difficult with climate changes and their impact on crop health and pest distribution, as accurate forecasting requires the integration of multi-source data such as soil properties, weather, crop production, and soil moisture measured via satellites such as SMAP. Therefore, there is a need to develop an intelligent system based on deep learning and

machine learning to analyze this diverse data, provide fast and accurate forecasts of diseases and pests, determine the best crops and the most appropriate planting and harvesting times, thereby enhancing agricultural productivity and reducing economic and environmental risks for farmers.

5. Motivation and Objectives

This project aims to empower farmers by enhancing their agricultural knowledge and enabling them to make informed decisions regarding optimal crop selection and timing. The objectives are to increase crop yield, improve crop quality, and provide farmers with timely, actionable information. The system will deliver monthly notifications for appropriate planting periods, alerts when crops are ready for harvest, and updates on crop health and disease status. Furthermore, it seeks to assist farmers in efficient water management, recommend suitable fertilizers and soil treatments, and promote sustainable farming practices to ensure long-term productivity and environmental conservation.

The main objectives of the proposal are as follows:

1- To integrate various data sources, including soil moisture from SMAP satellites, weather conditions, soil properties, and crop production history, to enhance the accuracy of crop predictions and recommendations.
2- To use AI-driven models for the early detection of agricultural pests and diseases, allowing farmers to take timely action and prevent potential damage to crops.
3- To offer an Android-based application that delivers real-time notifications and recommendations, ensuring that farmers have up-to-date information about crop health, pest threats, and the best practices for improving their farm productivity.

6. Methodology

The methodology used for developing a crop recommendation system involves several key stages, beginning with the collection of diverse datasets followed by the preprocessing, model training, and deployment phases. The primary goal of this methodology is to ensure the system is capable of providing accurate and context-specific crop recommendations based on various factors such as soil conditions, climate, and environmental factors. The approach incorporates advanced ML and DL models, leveraging Internet of Things (IoT) data, soil moisture readings, and climatic data to provide insights for agricultural decisions. This methodology is structured into four main stages: data collection, dataset splitting and training, new dataset collection, and deployment through cloud integration and Android application for end-user recommendations as shown in Figure 1.

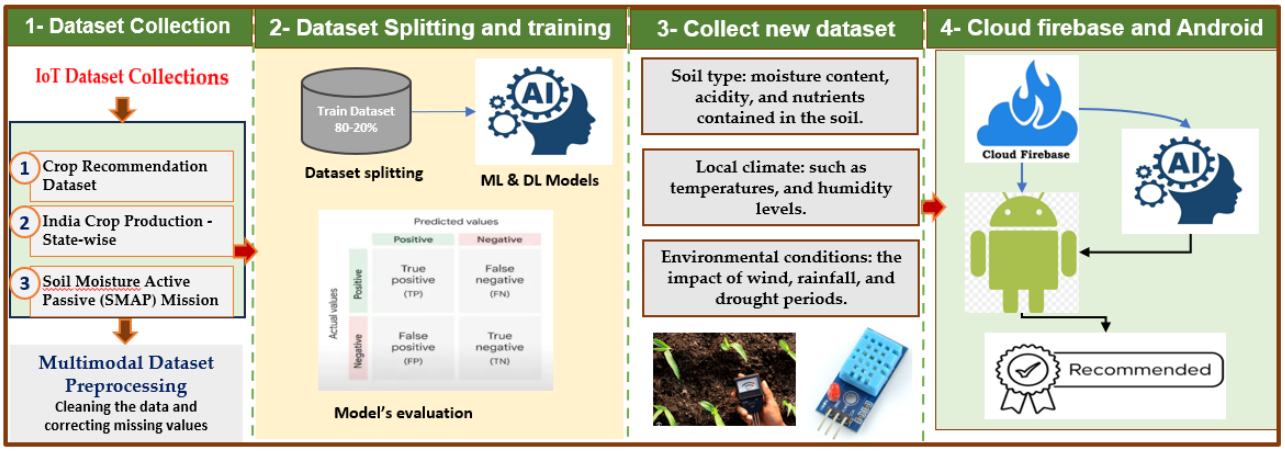
1- Dataset Collection
IoT Dataset Collections
1 Crop Recommendation Dataset
2 India Crop Production - State-wise
3 Soil Moisture Active Passive (SMAP) Mission
Multimodal Dataset Preprocessing
Cleaning the data and correcting missing values
2- Dataset Splitting and training
Train Dataset 80-20%
Dataset splitting
ML & DL Models
Predicted values
Positive
Negative
Actual values
True positive (TP)
False negative (FN)
False positive (FP)
True negative (TN)
Model's evaluation
3- Collect new dataset
Soil type: moisture content, acidity, and nutrients contained in the soil.
Local climate: such as temperatures, and humidity levels.
Environmental conditions: the impact of wind, rainfall, and drought periods.
4- Cloud firebase and Android
Cloud Firebase
Recommended

Figure 1: The proposed AI-driven framework for crop predication.

6.1. Dataset description

A vast fraction of the population of India considers agriculture as its primary occupation. The production of crops plays an important role in our country. Bad quality crop production is often due to either excessive use of fertilizer or using not enough fertilizer. The proposed system of IoT and ML is enabled for soil testing using the sensors, is based on measuring and observing soil parameters. This system lowers the probability of soil degradation and helps maintain crop health.[15]

6.1.1. Crop Recommendation Dataset

A dataset that recommends crops usually has information about the soil and environment, such as the levels of nitrogen (N), phosphorus (P), and potassium (K), the pH, temperature, humidity, and rainfall [16]. This dataset is used to figure out which crop will grow best in a certain area. Farmers can use these datasets for machine learning to make better decisions about how to manage their crops and resources, which will help them grow more crops and make more money in the long run. The output is a categorical variable that shows what crop is best. Important parts and features Soil nutrients: It has the right amounts of important macronutrients like Nitrogen (N), Phosphorus (P), and Potassium (K). Soil conditions: This includes tests for the soil's acidity or alkalinity (pH) and the amount of organic matter it has. Environmental factors: This includes information about the temperature, humidity, and amount of rain. Variable to aim for: The last column is a label that tells you what type of crop would grow best in the given conditions.

6.1.2. India Crop Production - State-wise

India's crop production varies significantly by state, with leading producers specializing in specific crops: Uttar Pradesh is the largest producer of food grains and sugarcane; Punjab is known as the "Food Bowl" for wheat and rice; Maharashtra leads in sugarcane and cotton; and West Bengal is a top producer of rice and jute. Other major crops and their key producing states

include soybean (Madhya Pradesh), coffee (Karnataka), tea (Assam, West Bengal, Kerala), and pulses (Madhya Pradesh, Maharashtra, Rajasthan) [17].

6.1.3. The Soil Moisture Active Passive (SMAP) Mission

The Soil Moisture Active Passive (SMAP) mission is one of the first Earth observation satellites being developed by NASA in response to the National Research Council's Decadal Survey. SMAP will make global measurements of the soil moisture present at the Earth's land surface and will distinguish frozen from thawed land surfaces. Direct observations of soil moisture and freeze/thaw state from space will allow significantly improved estimates of water, energy, and carbon transfers between the land and the atmosphere [18].

6.1.4. Collecting a new dataset

In this context, a crop recommendation dataset is being collected with the aim of providing a model based on multiple pieces of information such as soil type, climate, previously grown crops, and weather variations. The system aims to provide farmers with recommendations on which crops to plant in specific areas based on environmental and productivity factors.

- **Variables to be Collected**

When collecting data to build a crop recommendation model, several variables that directly affect the choice of the appropriate crop must be considered, such as:

1. Soil type: moisture content, acidity, and nutrients contained in the soil.
2. Local climate: such as temperatures, humidity levels, and the duration of seasons.
3. Environmental conditions: the impact of wind, rainfall, and drought periods.

- **Data Collection Methods**

Ground sensors: which can be installed in the fields to obtain real-time data on environmental characteristics.

6.2. Dataset Preprocessing and Feature Selection

The initial data was processed using the traditional study of collecting data from more than one study and using satellites, and then data processing was developed using ML and a system was created to facilitate data processing.

6.3. AI models

Artificial intelligence models and trains the recommendation system using pre-processed data and machine learning algorithms. To train the model, the data set must be divided into two parts, that is, the training data set and the test data set. Using ML algorithms on the training dataset we will train and build the model.[20]

6.3.1. ML

ML can stand for both machine learning and milliliters. Machine learning is a type of artificial intelligence where systems learn from data to make predictions or decisions without explicitly

programming them. A milliliter (ML) is a unit of volume, equal to one thousandth of a liter, used in measurements of liquids.

6.3.2. Convolutional Neural Network (CNN) and Transfer learning

CNNs are a class of deep neural networks that have proven to be highly effective in the domain of image processing and computer vision. CNNs are specifically designed to recognize patterns and structures in data, particularly visual data such as images and videos, by mimicking the way humans process visual information. A key feature of CNNs is their ability to automatically and adaptively learn spatial hierarchies of features through the convolution operation.

CNNs consist of multiple layers, including convolutional layers, pooling layers, and fully connected layers. The convolutional layers apply filters (or kernels) to input images, extracting low-level features such as edges, textures, and shapes. These features are then progressively combined through deeper layers to form high-level representations, enabling the network to identify more complex patterns, such as objects or faces. Pooling layers, typically using operations such as max pooling, are employed to reduce the spatial dimensions of the input while retaining essential features, which helps in reducing computation and preventing overfitting [21].

The final layers of a CNN, often fully connected layers, are used for classification tasks, where the learned features are used to predict the class of an image. CNNs have been extensively used in applications such as image classification, object detection, and facial recognition, among others, due to their ability to process and learn from large amounts of data in an efficient and hierarchical manner.

Transfer learning is a machine learning technique that allows a model developed for a specific task to be reused on another related task. The core idea behind transfer learning is to take advantage of the knowledge learned by a model on a large-scale dataset and apply it to a new task with limited data. This is especially useful in domains where collecting a large labeled dataset is expensive or time-consuming [22].

In the context of CNNs, transfer learning typically involves using a pre-trained network (usually trained on a large, generic dataset such as ImageNet) and fine-tuning it for a specific application. Instead of training a CNN from scratch, which requires substantial computational resources and large datasets, a pre-trained model can be leveraged to extract general features from images. These features can then be adapted to a new domain with a smaller dataset, leading to faster convergence and improved performance.

There are two primary approaches in transfer learning: feature extraction and fine-tuning [23]. In feature extraction, the pre-trained CNN is used as a fixed feature extractor, where the output of the convolutional layers is fed into a new classifier. In fine-tuning, the weights of the pre-

trained model are adjusted slightly by training the model on the new dataset, allowing the model to adapt its learned features to the new task.

6.3.3. Federated Learning

Federated learning unites multiple participants to generate a shareable global optimal model and keeps privacy-sensitive data on local devices, which may bring great hope to us for solving the problems of decentralized data and privacy protection [24], [25].

6.4. Android Application for Agriculture Recommendation

The final stage involves integrating the trained model with Cloud Firebase, which allows real-time data storage and accessibility for users. The system is connected to an Android application, providing an interface where farmers can access crop recommendations and receive real-time updates based on the current conditions. The recommendations provided by the system are marked as Recommended, based on the model's predictions and the current environmental conditions [26], [27].

7. Expected outputs

Expected results from the study, since we used artificial intelligence, the results will be as follows

- The work of the Android system to facilitate the work of farmers.
- Know the farmer when the right planting time.
- Know the farmer when it's time to collect crops.
- Farmers know for all lands what crops they can grow.

8. Conclusion

This study demonstrated the potential of machine learning and deep learning techniques in predicting crop diseases and pest outbreaks using multisource environmental datasets. Despite the strong performance of many reviewed models, the literature revealed several limitations, including insufficient high-quality data, limited regional generalization, and the need for better integration of remote-sensing technologies such as SMAP. The proposed system addresses these challenges by combining diverse datasets, improved preprocessing, and advanced AI models, supported by an Android application that delivers real-time planting, harvesting, and crop status notifications to farmers. These tools enhance decision-making, increase productivity, and support sustainable agricultural practices. Overall, AI-driven agriculture stands as a promising direction for improving crop health monitoring and yield prediction. and refining environmental modeling to further strengthen prediction accuracy.

References

[1] S. D. Tanmay Banavlikar, Aqsa Mahir, Mayuresh Budukh, “CROP RECOMMENDATION SYSTEM USING NEURAL NETWORKS,” *IRJET*, 2018.

[2] A. D. Rohit Kumar Rajak, Ankit Pawar, Mitalee Pendke, Pooja Shinde, Suresh Rathod, "Crop Recommendation System to Maximize Crop Yield using Machine Learning Technique," *Int. Res. J. Eng.*, 2017.

[3] S. R. M. B.K. Wani, R.N. Kaspate, "Growth and instability in area, production and productivity of major cash crops in India," *Eco. Env. Cons*, 2024, doi: http://doi.org/10.53550/EEC.2024.v30i02.073.

[4] I. K. Christine Musanase, Anthony Vodacek, Damien Hanyurwimfura, Alfred Uwitonze, "Data-Driven Analysis and Machine Learning-Based Crop and Fertilizer Recommendation System for Revolutionizing Farming Practices," *Agric. (volume 13, Issue 11, Artic. number 2141)*, 2023.

[5] M. K. Senapaty, A. Ray, and N. Padhy, "A Decision Support System for Crop Recommendation Using Machine Learning Classification Algorithms," *Agriculture*, vol. 14, no. 8, p. 1256, Jul. 2024, doi: 10.3390/agriculture14081256.

[6] U. R. N. Dagada Madhu Sudhan ReddyAnd, "A Comparative Analysis of Machine Learning Models for Crop Recommendation in India," 2023, doi: 10.18280/ria.370430.

[7] A. Srinidhi, V. Jahnavi, and M. Dholvan, "A Machine Learning and Deep Learning-Based Web Application for Crop and Fertilizer Recommendation and Crop Disease Prediction," 2024, pp. 389–402. doi: 10.1007/978-981-99-8451-0_33.

[8] S. K. Upadhyay, H. Upadhyay, H. Tripathi, K. P. Rawat, and L. Sharma, "Smart Crop Recommendation for Modern Farming using Machine Learning," in *2024 1st International Conference on Advanced Computing and Emerging Technologies (ACET)*, IEEE, Aug. 2024, pp. 1–7. doi: 10.1109/ACET61898.2024.10730582.

[9] H. Mavi, S. K. Upadhyay, N. Srivastava, R. Sharma, and R. Bhargava, "Crop Recommendation System Based on Soil Quality and Environmental Factors Using Machine Learning," in *2024 International Conference on Emerging Innovations and Advanced Computing (INNOCOMP)*, IEEE, May 2024, pp. 507–512. doi: 10.1109/INNOCOMP63224.2024.00089.

[10] S. Bhatnagar, N. Lakshmi, A. A, and V. S. M. Y, "Development of a Crop Recommendation System Through the Use of Various Machine Learning Algorithms," in *2024 Second International Conference on Emerging Trends in Information Technology and Engineering (ICETITE)*, IEEE, Feb. 2024, pp. 1–6. doi: 10.1109/ic-ETITE58242.2024.10493758.

[11] M. Pan, X. Cai, N. W. Chaney, D. Entekhabi, and E. F. Wood, "An initial assessment of SMAP soil moisture retrievals using high-resolution model simulations and in situ observations," *Geophys. Res. Lett.*, vol. 43, no. 18, pp. 9662–9668, Sep. 2016, doi: 10.1002/2016GL069964.

[12] G. Singh, N. Das, R. Panda, B. Mohanty, D. Entekhabi, and B. Bhattacharya, "Soil Moisture Retrieval Using SMAP L-Band Radiometer and RISAT-1 C-Band SAR Data in the Paddy Dominated Tropical Region of India," *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.*, vol. 14, pp. 10644–10664, 2021, doi: 10.1109/JSTARS.2021.3117273.

[13] M. A. and T. D. Woldekidan Gudelo‹, "Boosting Agricultural Outcomes with a Hybrid Ensemble Model for Crop Prediction under Climatic and Soil Variability," 2025, doi: //cybersecurityjournal.info/ 202.

[14] and K. K. C. Ryoko Araki, Bryn E. Morgan, Hilary K. McMillan, "Nonlinear Soil Moisture Loss Function Reveals Vegetation Responses to Water Availability," 2025, doi: //doi.org/10.1029/2024GL111403.

[15] D. Gosai, C. Raval, R. Nayak, H. Jayswal, and A. Patel, "Crop Recommendation System using Machine Learning," *Int. J. Sci. Res. Comput. Sci. Eng. Inf. Technol.*, pp. 558–569, Jun. 2021, doi: 10.32628/CSEIT2173129.

[16] V. K. PATEL, "Crop Recommendation dataset," 2024. doi: 10.21227/12nr-fe03.

[17] G. Sravanthi and N. R. Moparthi, "An efficient IoT based crop disease prediction and crop recommendation for precision agriculture," *Cluster Comput.*, vol. 27, no. 5, pp. 5755–5782, Aug. 2024, doi: 10.1007/s10586-023-04246-w.

[18] D. Entekhabi *et al.*, "The Soil Moisture Active Passive (SMAP) Mission," *Proc. IEEE*, vol. 98, no. 5, pp. 704–716, May 2010, doi: 10.1109/JPROC.2010.2043918.

[19] S. Pudumalar, E. Ramanujam, R. H. Rajashree, C. Kavya, T. Kiruthika, and J. Nisha, "Crop recommendation system for precision agriculture," in *2016 Eighth International Conference on Advanced Computing (ICoAC)*, IEEE, Jan. 2017, pp. 32–36. doi: 10.1109/ICoAC.2017.7951740.

[20] S. M. PANDE, P. K. RAMESH, A. ANMOL, B. R. AISHWARYA, K. ROHILLA, and K. SHAURYA, "Crop Recommender System Using Machine Learning Approach," in *2021 5th International Conference on Computing Methodologies and Communication (ICCMC)*, IEEE, Apr. 2021, pp. 1066–1071. doi: 10.1109/ICCMC51019.2021.9418351.

[21] A. M. Al-Hejri, R. M. Al-Tam, M. Fazea, A. H. Sable, S. Lee, and M. A. Al-antari, "ETECADx: Ensemble Self-Attention Transformer Encoder for Breast Cancer Diagnosis Using Full-Field Digital X-ray Breast Images," *Diagnostics*, vol. 13, no. 1, p. 89, Dec. 2022, doi: 10.3390/diagnostics13010089.

[22] M. R. Al-Maamari, R. Ramteke, A. M. Al-Hejri, and S. S. Alshamrani, "Integrating CNN and transformer architectures for superior Arabic printed and handwriting characters classification," *Sci. Rep.*, vol. 15, no. 1, p. 29936, Aug. 2025, doi: 10.1038/s41598-025-12045-z.

[23] A. M. Al-Hejri, R. M. Al-Tam, A. H. Sable, B. Almuhaya, S. S. Alshamrani, and K. M. Alshmrany, "A hybrid vision transformer with ensemble CNN framework for cervical cancer diagnosis," *BMC Med. Inform. Decis. Mak.*, vol. 25, no. 1, p. 411, Nov. 2025, doi: 10.1186/s12911-025-03250-x.

[24] A. M. Al-Hejri *et al.*, "A hybrid explainable federated-based vision transformer framework for breast cancer prediction via risk factors," *Sci. Rep.*, vol. 15, no. 1, p. 18453, May 2025, doi: 10.1038/s41598-025-96527-0.

[25] B. Yu, W. Mao, Y. Lv, C. Zhang, and Y. Xie, "A survey on federated learning in data mining," *WIREs Data Min. Knowl. Discov.*, vol. 12, no. 1, Jan. 2022, doi: 10.1002/widm.1443.

[26] R. N. Athirah, C. Y. N. Norasma, and M. R. Ismail, "Development of an Android Application for Smart Farming in Crop Management," *IOP Conf. Ser. Earth Environ. Sci.*, vol. 540, no. 1, p. 012074, Jul. 2020, doi: 10.1088/1755-1315/540/1/012074.

[27] A. M. Al-Hejri, N. M. Al-Zidi, M. Tawfik, A. Albakhrani, and A. H. Sable, "A Facilitation System for Arabic Foreigners in India Using the Web and Android System," in *2022 8th International Conference on Advanced Computing and Communication Systems (ICACCS)*, IEEE, Mar. 2022, pp. 239–245. doi: 10.1109/ICACCS54159.2022.9785022.